

Carstens, Carola; Rittberger, Marc

Einsatz von Semantic Web-Technologien am Informationszentrum Bildung

formal überarbeitete Version der Originalveröffentlichung in:

formally revised edition of the original source in:

Ockenfeld, Marlies [Hrsg.]: *Semantic web & linked data*. Frankfurt, Main : DGI 2010, S. 145-154



Bitte verwenden Sie in der Quellenangabe folgende URN oder DOI /

Please use the following URN or DOI for reference:

urn:nbn:de:0111-dipf-37902

10.25657/02:3790

<https://nbn-resolving.org/urn:nbn:de:0111-dipf-37902>

<https://doi.org/10.25657/02:3790>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

DIPF | Leibniz-Institut für
Bildungsforschung und Bildungsinformation
Frankfurter Forschungsbibliothek
publikationen@dipf.de
www.dipfdocs.de

Mitglied der


Leibniz-Gemeinschaft

Carola Carstens, Marc Rittberger

Deutsches Institut für Internationale Pädagogische Forschung, Informationszentrum Bildung

Einsatz von Semantic Web-Technologien am Informationszentrum Bildung

Abstract

Dieser Beitrag gibt einen Überblick über die Aktivitäten des Informationszentrums Bildung zum Einsatz von Semantic Web-Technologien. Schwerpunkte liegen hierbei auf der Beschreibung des Aufbaus einer Ontologie für die Domäne der Bildungsforschung, sowie auf der Darstellung, wie sie zum Zweck der Query Expansion in einem domänenspezifischen Retrievalsystem genutzt werden kann.

Aktuelle Semantic Web-Entwicklungen

Vor knapp 10 Jahren beschrieben [Berners-Lee et al. 2001] ihre Vision eines Semantic Web als Erweiterung des bestehenden Webs. Diese Erweiterung aus semantischen Annotationen soll im Gegensatz zum herkömmlichen Web nicht nur von Menschen interpretierbar sein, sondern auch von Maschinen.

Zu diesem Zweck wurden Standards entwickelt, mit denen Daten und Dokumenteninhalte semantisch beschrieben werden können. Mithilfe der Semantic Web-Standards RDFS¹ und OWL² lassen sich Ontologieschemata modellieren, die das Vokabular zur semantischen Beschreibung von Daten und Dokumenteninhalten im RDF³-Format definieren. Ein solches Schema besteht aus für eine Domäne spezifischen Klassen und Relationen. In einem Ontologieschema zur Beschreibung von Forschungskontexten lässt sich bspw. festlegen, dass Personen verknüpft sind mit den Organisationen, in denen sie arbeiten, sowie mit gegebenenfalls vorhandenen Unterorganisationen. Diese Organisationen führen jeweils Projekte durch, an denen wiederum Personen arbeiten. Dieses Schema lässt sich anschließend nutzen, um die Beziehungen zwischen einzelnen Instanzen von Personen und Organisationen zu beschreiben.

Ursprünglich wurden diese Standards entwickelt, um unstrukturierte Informationen im herkömmlichen Web semantisch annotieren und in maschinenlesbarer Form repräsentieren zu können. Auf diese Weise sollten eigenständig handelnde Agentenprogramme in die Lage versetzt werden, Daten aus verschiedenen Informationsquellen zusammenzuführen, automatisch Schlüsse zu ziehen und Nutzer bei Entscheidungsprozessen zu unterstützen.

Allerdings erfordert die semantische Annotation von Informationen im Web einen Arbeitsaufwand, der für die Dateneigner zunächst keinen direkten Mehrwert bietet. Dieser wird erst deutlich, wenn sie die eigenen Daten mit anderen Datenquellen verknüpfen können, um Datenquellen übergreifende Mehrwertdienste zu implementieren. Damit sich derartige Dienste realisieren lassen, muss das Semantic Web allerdings erst eine gewisse Menge an Daten umfassen.

Neben herkömmlichen Webseiten stellen strukturierte Daten eine vielversprechende Datenquelle für das Semantic Web dar. Diese werden bisher zumeist in relationalen Datenbanken gespeichert und lassen sich einfacher semantisch beschreiben als textuelle Webseiteninhalte. Während der ursprüngliche Fokus des Semantic Web auf der Annotation von Webseiteninhalten lag, ist daher inzwischen die Publikation von diesen bereits in strukturierter Form vorliegenden Daten in das Zentrum des Interesses gerückt. Sie stellen eine

¹ <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>

² <http://www.w3.org/TR/2004/REC-owl-features-20040210/>

³ <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>

wichtige Ressource für das Semantic Web dar, deren Aufbereitung in Semantic Web-Standards in der Linked Data-Initiative [Berners-Lee 2006] vorangetrieben wird.

Hierbei steht das Ziel im Fokus, möglichst viele Datenbestände in Semantic Web-Standards zu überführen, sie zu publizieren, und Anknüpfungspunkte zu anderen Datenbeständen zu definieren. Dies kann z.B. durch die Definition von Äquivalenzbeziehungen erreicht werden, doch auch jede Art anderer semantischer Relationen ist denkbar. Auf diese Weise soll schließlich ein Netz verknüpfter Datenbestände entstehen, das „Web of Linked Data“, das mittlerweile häufig synonym zum Semantic Web benutzt wird.

Diese Initiative hat inzwischen auch Einzug in den deutschen Bibliotheksbereich gefunden. Schließlich verfügen Bibliotheken über normierte und gut erschlossene Datenbestände, sind Spezialisten in der dokumentarischen Beschreibung von Daten und verfügen somit über ein großes Potenzial, zur Realisierung des Semantic Web beizutragen. So hat bspw. die Deutsche Nationalbibliothek bereits einen Linked Data-Service eingerichtet und strebt an, alle nationalbibliografischen Daten inklusive aller Normdaten als Linked Data zu publizieren [Kett 2009]. Ähnliche Initiativen aus dem Bibliotheksbereich sind z.B. von der Schwedischen Nationalbibliothek bekannt [Söderbäck 2009], sowie von der Deutschen Zentralbibliothek für Wirtschaftswissenschaften [Neubert 2009].

Sobald einzelne Datenquellen mit Semantic Web-Standards annotiert sind, lassen sie sich mithilfe dieser Standards auch leicht miteinander verknüpfen, wobei semantische Beziehungen zwischen Daten aus den verschiedenen Quellen definiert werden können. So lässt sich bspw. definieren, dass alle Universitäten aus einem Datenbestand zu einer Unterklasse der Klasse *Organisation* in einem anderen Datenbestand gehören. Dies wird realisiert, indem die jeweils zur Annotation verwendeten Ontologien miteinander in Beziehung gesetzt werden, ohne die Daten selbst transformieren zu müssen. Dadurch wird es möglich, Datenquellen übergreifende Suchanfragen zu stellen und Schlussfolgerungen über Daten aus verschiedenen Quellen zu ziehen.

Die Abfrage semantisch annotierter Daten ist über Schnittstellen zur Faktensuche realisierbar. Auf der Basis der oben beschriebenen Ontologie zur Abbildung von Forschungskontexten kann ein Nutzer sich bspw. alle zu einer Organisation zugehörigen Mitarbeiter ausgeben lassen. Werden bei dieser Abfrage darüber hinaus auch Reasoning-Mechanismen angewandt, so ist das automatische Ziehen von Schlüssen über die Daten möglich. Im genannten Beispiel könnte bspw. abgeleitet werden, dass alle Mitarbeiter einer Unterorganisation auch der Oberorganisation angehören, sodass diese Mitarbeiter auch als Suchergebnis ausgegeben werden könnten.

Darüber hinaus lassen sich semantisch annotierte Daten nutzen, um bspw. semantische Navigationsstrukturen mit typisierten Links aufzubauen, die die semantischen Beziehungen zwischen Informationen näher beschreiben. So würde ein Link von einer Projektseite zu einer Organisationsseite bspw. annotiert werden mit der Bedeutung *wird ausgeführt von*. Auch zur Unterstützung von Textretrieval-Prozessen sind Ontologien einsetzbar, beispielsweise für die Query Expansion (Suchanfrageerweiterung) oder für die Generierung von Suchvorschlägen.

Der Fokus des Linked Data-Ansatzes liegt darauf, publizierte Datenbestände verschiedener Quellen miteinander semantisch zu verknüpfen und auf dieser Basis Mehrwertdienste wie die oben beschriebenen anzubieten. Beispiele hierfür sind Suchdienste, Reasoning-Mechanismen und Navigationsstrukturen, die eine integrierte Sicht auf verteilte Datenbestände ermöglichen. Darüber hinaus können die zu diesem Zweck entwickelten Semantic Web-Technologien allerdings auch für interne Zwecke von Nutzen sein. Hierzu zählen bspw. die Datenintegration und die Entwicklung von semantischen Suchdiensten auf der Basis der Datenbestände eines einzelnen Dateneigners [Carstens 2008].

Aktivitäten des IZB: Schwerpunkt Query Expansion

Auch im Informationszentrum Bildung (IZB) des Deutschen Instituts für Internationale Pädagogische Forschung (DIPF) kommen Semantic Web-Technologien zum Einsatz. Das IZB hat eine ontologische Wissensbasis für den Bereich der Bildungsforschung aufgebaut. Mithilfe von Semantic Web-Technologien wurden Daten aus verteilten Datenbeständen in dieser Ontologie semantisch integriert. Als ein Beispiel für den Mehrwert der Ontologie wurde zunächst ihr Potenzial für den Anwendungsfall der Query Expansion evaluiert.

Der vorliegende Beitrag beschreibt den Aufbau dieser Ontologie und gibt einen Überblick über Mehrwerte der ontologiebasierten Query Expansion. Dabei werden auch Probleme und Grenzen dargestellt, die beim Einsatz der semantischen Technologien identifiziert wurden. Abschließend wird ein Ausblick auf weitere Semantic Web-Aktivitäten des IZB gegeben, die sich aus der Nachnutzung und dem Ausbau der bisher geschaffenen Strukturen ergeben werden.

Aufbau einer ontologischen Wissensbasis für die Erziehungswissenschaft

Die im IZB modellierte Ontologie bildet Forschungskontexte aus der Domäne der Bildungsforschung ab. Ihre Struktur wurde allgemein und ausbaufähig definiert, aber auf den speziellen Anwendungsfall der Query Expansion fokussiert. Momentan dient sie als Quelle für die Generierung von Query Expansion-Termen in einem domänenspezifischen Retrieval-Prototypen, der auf einem Teilkorpus der FIS Bildung-Literaturdatenbank⁴ basiert.

Zu diesem Zweck wurde zunächst ein einfaches Ontologieschema modelliert, das aus den Hauptklassen *Projekt*, *Person*, *Konzept* und *Organisation* sowie deren Beziehungen untereinander besteht. Sämtliche Forschungsinhalte der Bildungsforschung werden hierbei als Konzepte aufgefasst. Dieses Schema wurde im eigens definierten Namensraum RESCON (*Research Context Ontology*) modelliert. Abbildung 1 gibt einen Überblick über die Hauptklassen und –relationen dieses Ontologieschemas.

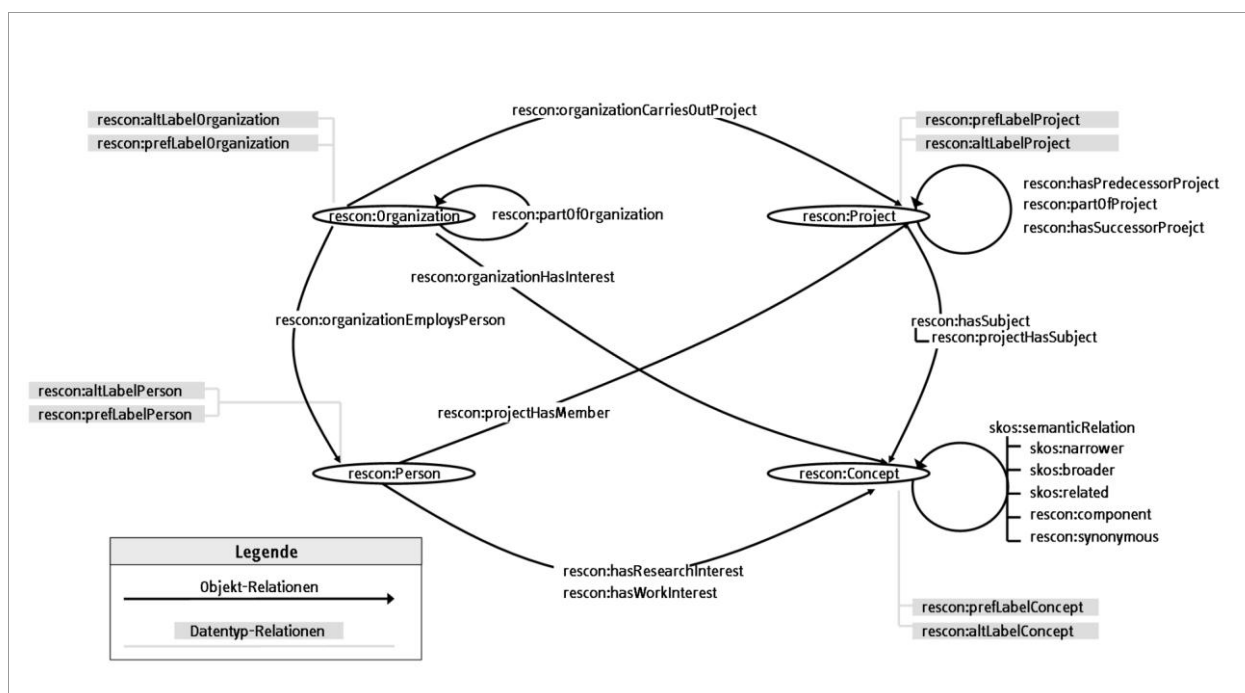


Abbildung 1: RESCON-Ontologieschema

Beim Aufbau der Ontologie wurde zudem Wert darauf gelegt, Anknüpfungspunkte an bereits bestehende Semantic Web-Vokabulare zu definieren. Auf diese Weise soll das Ziel des

⁴ http://www.fachportal-paedagogik.de/fis_bildung/index.html

Semantic Web und des Linked Data-Ansatzes unterstützt werden, publizierte Datenbestände leicht für andere Anwendungen interpretierbar zu machen. Für die Darstellung von Forschungskontexten wurden hierbei die Vokabulare FOAF, SKOS und SWRC als geeignete und integrierbare Ontologien identifiziert.

FOAF⁵ ermöglicht es, Personenprofile zu beschreiben, bspw. anhand des Namens (*foaf:name*), des Alters (*foaf:age*) oder anhand zwischenmenschlicher Beziehungen (*foaf:knows*). Die SWRC-Ontologie [Sure et al. 2005] dient hingegen explizit der Beschreibung von Forschungskontexten. So lassen sich bspw. Beziehungen zwischen den Klassen *swrc:University*, *swrc:FacultyMember* und *swrc:Project* abbilden. SKOS⁶ hingegen ist ein Standard für die Überführung von Vokabularen wie Thesauri in Semantic Web-Standards. Hiermit lassen sich bspw. typische Thesaurusrelationen wie *skos:synonymous*, *skos:broader* und *skos:narrower* zwischen Instanzen der Klasse *skos:Concept* definieren.

Das in Abbildung 1 dargestellte Ontologieschema wurde mit diesen externen Ontologien in Beziehung gesetzt. Zu diesem Zweck wurden die RESCON-Klassen als Unterklassen der jeweiligen externen Klassen definiert, bspw. *rescon:Person* als Unterklasse von *foaf:Person*. Auf diese Weise war es möglich, Relationen zwischen extern definierten Klassen im eigenen Namespace zu nutzen, während sich neu definierte Relationen nur auf die selbst definierten Unterklassen beziehen.

So wurden bspw. Relationen wie *skos:semanticRelation* und dessen Unterrelationen *skos:broader* und *skos:narrower* aus dem SKOS-Vokabular übernommen (siehe Abb. 1). Für den eigenen Anwendungsfall mussten die Relationen allerdings noch ergänzt werden. So wurden u. a. weitere Unterrelationen von *skos:semanticRelation* angelegt, die im SKOS-Vokabular nicht enthalten waren, wie zum Beispiel *rescon:synonymous*.

Ziel war es, möglichst viele Klassen und Relationen aus bestehenden Vokabularen wiederzuverwenden. Die in diesem Zuge getroffenen Modellierungsentscheidungen werden detailliert beschrieben in [Carstens 2009b].

Dieses Ontologieschema wurde anschließend mit Instanzdaten aus verschiedenen Datenquellen befüllt. Hierzu zählen Informationen über Personen aus einer Personendatenbank, die Forscher und Praktiker aus der Domäne verzeichnet, Informationen über bildungsbezogene Organisationen aus einer Organisationendatenbank⁷, sowie Informationen über innovative Projekte der Bildungsforschung aus einer Projektdatenbank⁸.

Darüber hinaus wurde ein im IZB für Dokumentationszwecke genutztes thesaurusartiges Schlagwortverzeichnis in die Ontologie integriert. Es umfasst typische Thesaurusrelationen wie Synonyme, Ober- und Unterbegriffe, sowie verwandte Begriffe. Diese Terme wurden als Instanzen der Klasse *skos:Concept* angelegt und in Anlehnung an Best Practices zur Überführung von Thesauri im SKOS-Standard abgebildet [Miles 2005; Assem 2006].

Dieses Vorgehen folgte dem pragmatischen Ansatz, die Ontologie nicht von Grund auf neu zu konzeptionalisieren, sondern Daten aus bestehenden Quellen wiederzuverwenden und semantisch zu integrieren. So konnten bspw. Informationen über Organisationen mit dem Thesaurusvokabular in Beziehung gesetzt werden. Da die Organisationendatenbank u. a. Forschungsschwerpunkte der gelisteten Organisationen verzeichnet, können diese Terme auf Konzepte des Thesaurus gemappt werden. So verzeichnet die Ontologie bspw. für die Organisation mit dem Label *Fachverband für Kunstpädagogik* die Relation *rescon:organizationHasInterest*, welche die Organisation mit dem Konzept *Ästhetische*

⁵ <http://xmlns.com/foaf/spec/20100101.html>

⁶ <http://www.w3.org/TR/skos-reference/>

⁷ <http://www.bildungsserver.de/institutionen.html>

⁸ <http://www.bildungsserver.de/innovationsportal/>

Erziehung verbindet. Dieses Konzept ist wiederum verknüpft mit dem Synonym *Ästhetisches Lernen*.

Insgesamt verzeichnet die auf diese Weise aufgebaute Ontologie mehr als 27.000 Konzepte, 5.500 Personen, 7.600 Organisationen und 500 Projekte. Zu den häufigsten Objektrelationen zwischen diesen Instanzen zählen *rescon:hasWorkInterests*, *skos:broader* und *rescon:synonymous*. Mithilfe von Reasoning-Mechanismen konnte die Ontologie darüber hinaus um weitere Relationen angereichert werden. So lassen sich bspw. über die Rule Language SWRL⁹ Zusammenhänge wie die folgenden definieren: *wenn eine Organisation zu einem Thema forscht, forscht sie implizit auch zum Oberthema*.

Zum Aufbau dieser Ontologie ist allerdings anzumerken, dass sie momentan keineswegs in der Lage ist, Forschungskontexte der Domäne vollständig abzubilden. Vielmehr diene sie bisher dem Zweck, als exemplarische ontologische Datenbasis zu dienen, um das Potenzial der einzelnen Typen von Daten und Zusammenhängen für den Anwendungsfall der Query Expansion zu untersuchen. Auf diesen Anwendungsfall soll daher im folgenden Abschnitt näher eingegangen werden.

Anwendungsfall: Ontologeeinsatz zur Query Expansion

Der Einsatz von Thesauri zum Zweck der Query Expansion hat bereits eine lange Tradition im Forschungsgebiet des Information Retrieval. Als korpusunabhängige Wissensstrukturen stellen Thesauri eine mögliche Quelle für die Generierung von Query Expansion-Termen dar [Efthimiadis 1996].

Durch die Aktivitäten im Semantic Web-Umfeld, die sich auch intensiv dem Thema der semantischen Suche widmen, sind die Forschungserkenntnisse zur wissensbasierten Query Expansion von ganz neuer Relevanz. Schließlich werden im Zuge der Semantic Web-Aktivitäten vielerorts Ontologien definiert und publiziert, die sich auch für die Query Expansion einsetzen lassen. Dies kann der wissensbasierten Query Expansion neuen Antrieb geben, mangelte es ihr in der Praxis doch häufig an umfassenden Vokabularen, deren Aufbau sehr ressourcenintensiv ist [Harman 1988; Greenberg 2001a].

[Efthimiadis 1996] unterscheidet zwei grundlegende Methoden zum Einsatz von Query Expansion-Termen. Demnach lassen sie sich einerseits für die automatische Query Expansion einsetzen, wobei Suchanfragen automatisch mit Termen aus einer Wissensquelle angereichert werden. Andererseits ist es auch möglich, dem Nutzer interaktive Query Expansion-Terme vorzuschlagen, die dieser nach eigenem Ermessen zur Reformulierung seiner Anfrage nutzen kann.

[Greenberg 2001b] stellt fest, dass sich Synonyme und Unterbegriffe für die automatische Query Expansion anböten, während bspw. verwandte Begriffe und Oberbegriffe eher für die interaktive Query Expansion geeignet seien. Bei den letzteren beiden Arten von Termen gehe ein positiver Effekt auf den Recall mit einem erheblichen Absinken der Precisionwerte einher. Daher sei es zu empfehlen, die Entscheidung über die Nutzung dieser Terme den Nutzern zu überlassen.

Auf der Basis dieser Erkenntnisse wird derzeit der Einsatz der RESCON-Ontologie zum Zwecke der Query Expansion in einem domänenspezifischen Retrievalsystem für die Bildungsforschung untersucht [Carstens 2009a].

Query Expansion für die FIS Bildung-Literaturdatenbank

Hierbei dient ein Auszug aus der FIS Bildung-Literaturdatenbank als Testkorpus, um die Effektivität von ontologiebasierten Query Expansion-Mechanismen exemplarisch für die

⁹ <http://www.daml.org/2003/11/swrl/>

Domäne zu untersuchen. Diese domänenspezifische bibliografische Referenzdatenbank umfasst mehr als 700.000 Metadatendokumente, größtenteils in deutscher Sprache.

In Anlehnung an [Greenberg 2001b] wird ermittelt, welche in der Ontologie abgebildeten semantischen Relationen sich für die automatische Query Expansion anbieten, und welche eher für die interaktive Query Expansion geeignet sind. Zu diesem Zweck wurde ein prototypisches Retrievalsystem erstellt, das eine interaktive ontologiebasierte Query Expansion-Komponente implementiert. Dieses wird derzeit im Rahmen von Nutzertests evaluiert. An dieser Stelle sollen einige beispielhafte Szenarien das Potenzial der Ontologie veranschaulichen.

Wie von [Greenberg 2001a] beschrieben, eignen sich Synonyme für die automatische Query Expansion. Demzufolge lässt sich eine Suche nach dem Term *Lernmotivation* bspw. automatisch mit dem Term *Lernbereitschaft* erweitern, der in der Ontologie als Synonym verzeichnet ist (siehe Abb. 2).

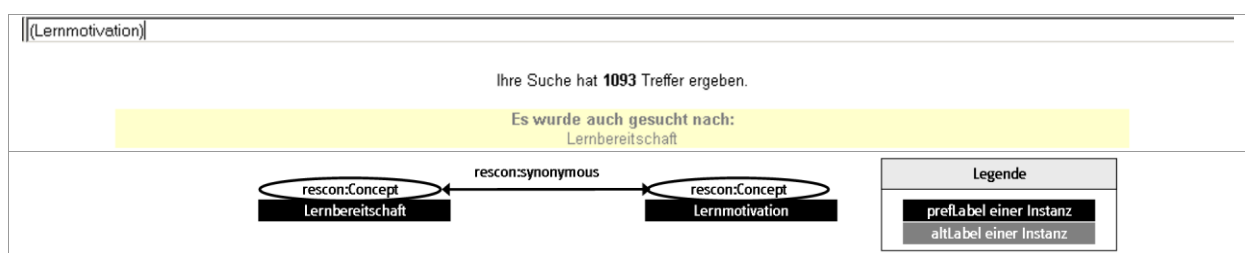


Abbildung 2: Beispiel für die automatische Query Expansion

Interaktive Expansionsterme werden dem Nutzer hingegen als Vorschläge angeboten, da sie vermutlich nur in bestimmten Situationen von Relevanz sind. So könnte die Suche nach den Termen *Sprachdiagnose AND Migranten* bspw. verwandte Terme wie *Migrantin* aus der Ontologie als interaktive Expansionsterme anzeigen. Darüber hinaus enthält die Ontologie auch Informationen über semantisch verknüpfte Projekte, die in diesem Kontext von Interesse sein können. Im Fall der oben beschriebenen Suche ließe sich bspw. eine Verknüpfung zu dem Projekt *FörMig* (*Förderung von Kindern und Jugendlichen mit Migrationshintergrund*) identifizieren. Wie in Abbildung 3 dargestellt, sind in der Ontologie die Themen *Sprachdiagnostik* und *Migranten* als Forschungsthemen des Projekts verzeichnet. Da der Term *Sprachdiagnostik* als Synonym zu *Sprachdiagnose* definiert ist, wird das Projekt *FörMig* als semantisch relationiertes Projekt identifiziert und dem Nutzer angezeigt. Dieser kann verschiedene Buttons nutzen, um einen oder mehrere der vorgeschlagenen Terme in seine Suchanfrage zu integrieren (siehe Abb. 3).

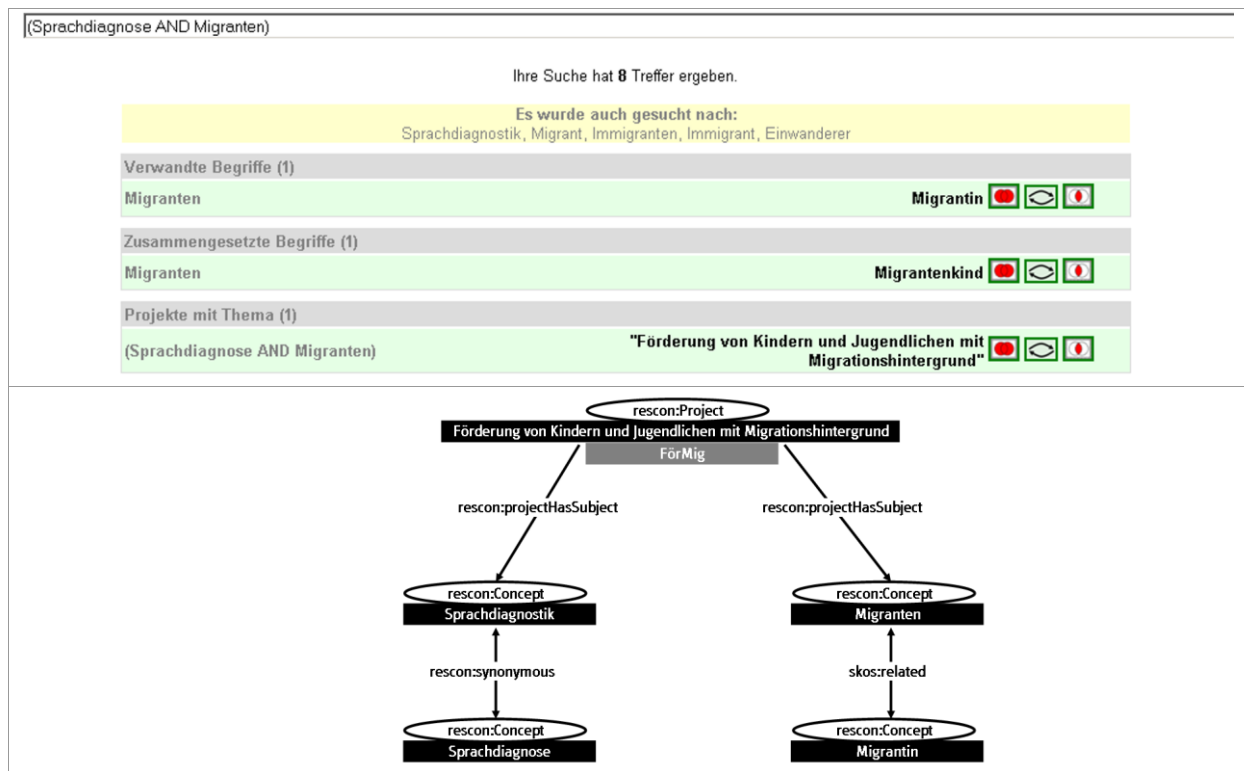


Abbildung 3: Beispiel für die interaktive Query Expansion

Darüber hinaus lässt sich die Ontologie auch zum Zweck der Disambiguierung nutzen. Sucht ein Nutzer bspw. nach dem Term *IGLU*, so lässt sich dieser Term auf zwei verschiedene Instanzen in der Ontologie mappen. Hierbei handelt es sich einerseits um das Projekt *IGLU* (*Internationale Grundschul-Lese-Untersuchung*), andererseits um das Konzept *Iglu* im Sinne von Behausung. Bei dieser Suchanfrage kann dem Nutzer bspw. der vollständige Projekttitel vorgeschlagen werden, mit dem er seine Suche ggf. präzisieren kann (siehe Abb. 4).

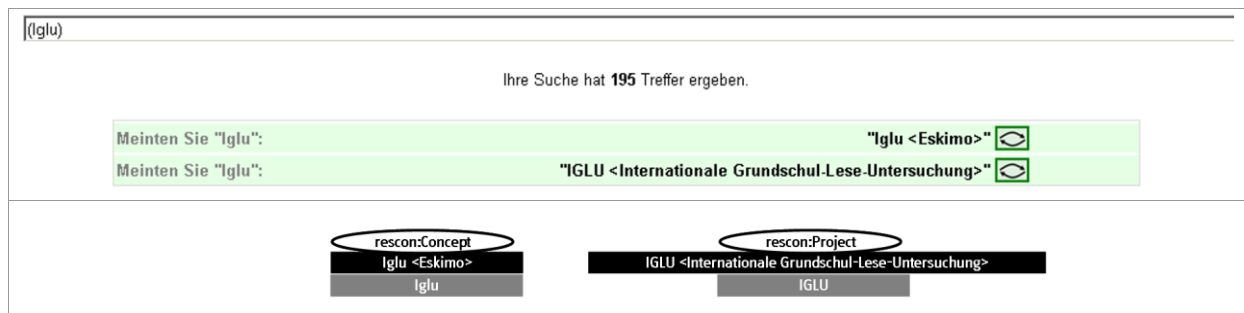


Abbildung 4: Beispiel für die Disambiguierung

Aktueller Stand der Semantic Web-Aktivitäten am IZB

Die Ontologie zur Bildungsforschung des IZB wird bisher hauptsächlich eingesetzt, um ihr Potenzial zum Zweck der Query Expansion zu untersuchen. Mit dem Retrievalsystem-Prototypen wurde hierzu eine Experimentierumgebung geschaffen. Die noch ausstehenden Ergebnisse der Query Expansion-Experimente werden das Potenzial für mögliche Anwendungen der Ontologie im Kontext der FIS-Bildung-Literaturdatenbank veranschaulichen.

Bei der Arbeit mit der Ontologie wurden darüber hinaus wertvolle Erfahrungen im Umgang mit Semantic Web-Technologien gesammelt. Auch momentane Einschränkungen bei der Nutzung dieser Technologien wurden dabei identifiziert. Während der nächste Abschnitt auf diese Erfahrungen eingeht, wird im darauffolgenden Abschnitt beschrieben, wie die erstellte

Ontologie in Zukunft als Ausgangspunkt für weitere Semantic Web-Aktivitäten des IZB dienen kann.

Erfahrungen aus dem Einsatz von Semantic Web-Technologien

Semantic Web-Technologien bieten die technologischen Grundlagen und die Syntax für die Definition von komplexen semantischen Beziehungen zwischen Daten. Dennoch können das Mapping bzw. die Integration verschiedener Datenbestände sehr ressourcenintensiv sein. Während die Ausdruckfähigkeit dieser Mappings durch die Nutzung von Semantic Web-Standards wie OWL sehr hoch ist, bleibt der Mapping-Aufwand ähnlich ressourcenintensiv wie bspw. bei der traditionellen Definition von Kookkurenzen zwischen Thesauri.

Der Einsatz der ontologischen Wissensbasis zur Query Expansion hat u. a. Grenzen der Performanz beim Einsatz von Semantic Web-Technologien verdeutlicht. Die Wissensbasis wurde im Prototyp als inferenziertes Modell in den Arbeitsspeicher geladen. Dabei steigt die erforderliche Ladezeit natürlich mit der Anzahl der enthaltenen Instanzdaten, die in der vorliegenden Wissensbasis sehr hoch ist. Bei einer weiteren Nutzung der Ontologie ließe sich die Performanz aber vermutlich durch den Einsatz eines Triple Stores verbessern, in dem die Daten jeweils in Aussagetripeln wie z.B. *Konzept A ist Oberbegriff von Konzept B* gespeichert sind. In diesem Fall muss beim Zugriff nicht mehr die gesamte instanziierte Ontologie in den Arbeitsspeicher geladen werden, sondern es können gezielte Anfragen an den performanzoptimierten Speicher gestellt werden.

Bei dem angewandten Verfahren der semi-automatischen Integration von Daten aus verschiedenen Ursprungsquellen in einer Ontologie stellt sich darüber hinaus die Frage, wie Aktualisierungsprozesse gestaltet werden können, welche die Ontologie möglichst automatisch mit aktualisierten Daten befüllen. Diese Frage stellt sich vor allem, falls die Ontologie weiter ausgebaut und genutzt werden soll, während die Ursprungsdaten in den Originaldatenquellen, bspw. relationalen Datenbanken, verbleiben. Sie ist daher strategischer Natur und leitet somit direkt über zum Ausblick auf weitere geplante Semantic Web-Aktivitäten des IZB.

Ausblick auf weitere Semantic Web-Aktivitäten

Das vorgestellte Ontologieschema lässt sich weiterhin zur Abbildung von Forschungskontexten nutzen und bietet auch Anknüpfungspunkte für die Anbindung externer Datenquellen. Diese Datenstruktur kann somit als Basis für die Implementierung von Datenquellen übergreifenden Informationsdiensten dienen.

Auch die Publikation eines Auszugs der Ontologie als Linked Data ist möglich. Vor der Publikation soll die für Evaluationszwecke aufgebaute Ontologie allerdings auf einen dokumentarisch geprüften Kernteil reduziert werden. Dies betrifft sowohl die dort verzeichneten domänenspezifischen Konzeptterme als auch die momentan große Anzahl an Organisationen-, Personen- und Projektinstanzen.

Mit seiner dokumentarischen Expertise in der Domäne der Bildungsforschung kann das IZB einen fachspezifischen Beitrag zum Semantic Web leisten. So pflegt das IZB bspw. einen Kernwortbestand für die Domäne der Bildungsforschung, der einen Auszug des anfangs erwähnten Thesaurus darstellt. Dieser Kernwortbestand wird laufend aktualisiert, definiert die bedeutendsten Begrifflichkeiten der Domäne und enthält teilweise auch semantische Relationen zwischen den Begriffen. Es bietet sich daher an, diesen geprüften Datenbestand als Linked Data zu publizieren, der als Referenzvokabular für die Domäne der Bildungsforschung dienen kann.

Darüber hinaus verfügen die Dokumentationsdienste des IZB über die Kompetenz, Kerninstitutionen, -projekte und -personen für die Domäne zu identifizieren. Diese können als Ausgangsbasis dienen, um strukturierte Zusatzinformationen aus externen Quellen zu

ermitteln, welche über die Ontologie mit den eigenen Daten in Beziehung gesetzt werden können. So lassen sich bspw. Autorennamen der Domäne mit in der Personennormdatei (PND) der DNB verzeichneten Personen mappen. Auf diese Weise können Personen bspw. mit Zusatzinformationen wie ihrer Organisation und ihrem Beruf aus der PND in Beziehung gesetzt werden.

Eine weitere Möglichkeit zur Verlinkung stellt die Datensammlung DBpedia dar, welche strukturierte Informationen aus Wikipedia in Semantic Web-Standards bereitstellt [Bizer et al. 2009]. Diese Quelle kann dazu dienen, strukturierte Informationen über Institutionen zu ermitteln. Die auf diese Weise zusammengeführten Zusatzinformationen lassen sich dann auch in Informationsdienste des IZB einbinden.

Darüber hinaus können evtl. noch weitere Datenquellen identifiziert werden, mit denen sich die Daten des IZB verlinken lassen. Zunächst müssen allerdings der Abdeckungsgrad und die Qualität der Daten in den zu verlinkenden Quellen geprüft werden. Voraussetzung ist lediglich, dass diese als Linked Data publiziert sind. Die Verlinkung und semantische Integration von Daten erlaubt dann die Implementierung Datenquellen übergreifender Dienste für die Domäne der Bildungsforschung.

In diesem Artikel wurde bereits dargestellt, wie sich Daten aus einer Forschungskontextontologie zur Query Expansion nutzen lassen. Darüber hinaus lassen sich Mechanismen zur Faktensuche, zum semantischen Reasoning und zum facettierten Browsing implementieren. Hierbei können insbesondere durch die Integration externer Datenquellen im Sinne des Linked Data-Ansatzes Mehrwerte generiert werden.

Je mehr Daten im Semantic Web verfügbar sind, desto attraktiver wird die semantische Verknüpfung von Datenbeständen werden. Hiermit wird auch der Mehrwert steigen, der sich aus der Implementierung der beschriebenen Datenquellen übergreifenden Dienste ergeben wird.

- Assem, M. v., Malaisé, V., Miles, A., Schreiber, G. (2006): A Method to Convert Thesauri to SKOS. In: Sure, Y., Domingue, J. (Hrsg.) 3rd European Semantic Web Conference 2006, Budva, Montenegro. LNCS Volume 4011, S. 95-109. Springer-Verlag: Heidelberg.
- Berners-Lee, T. (2006): Linked Data. <http://www.w3.org/DesignIssues/LinkedData.html>. (letzter Zugriff: 29.07.2010).
- Berners-Lee, T., Hendler, J., Lassila, O. (2001): The Semantic Web. *Scientific American*, 284(5), S. 34-43.
- Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S. (2009): DBpedia – A Crystallization Point for the Web of Data. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3), S. 154-164.
- Carstens, C. (2008): Semantic Web Applications in- and outside the Semantic Web. In: Auer, S., Schaffert, S., Pellegrini, T. (Hrsg.) I-SEMANTICS 2008, Graz, Österreich, S.25-33. J.UCS: Graz.
- Carstens, C. (2009a): Effects of Using a Research Context Ontology for Query Expansion. In: Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvönen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E. (Hrsg.) 6th European Semantic Web Conference 2009, Heraklion, Kreta, Griechenland. LNCS Volume 5554, S. 919-923. Springer-Verlag: Berlin, Heidelberg.
- Carstens, C. (2009b): Integrated Ontologies for the Semantic Web: Experiences from Modelling a Research Context Ontology. In: Paschke, A., Weigand, H., Behrendt, W., Tochtermann, K., Pellegrini, T. (Hrsg.) I-SEMANTICS 2009, Graz, Österreich, S. 630-637. Verlag der TU Graz: Graz.
- Efthimiadis, E.N. (1996): Query Expansion. In: William, M.E. (Hrsg.): *Annual Review of Information Systems and Technology*, Volume 31, S. 121-187. Information Today: Medford, New Jersey.
- Greenberg, J. (2001a): Automatic Query Expansion via Lexical-Semantic Relationships. *Journal of the American Society for Information Science and Technology*, 52(5), S. 402-415.
- Greenberg, J. (2001b): Optimal Query Expansion (QE) Processing Methods with Semantically Encoded Structured Thesauri Terminology. *Journal of the American Society for Information Science and Technology*, 52(6), S. 487-498.
- Harman, D. (1988): Towards Interactive Query Expansion. 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Grenoble, Frankreich, S. 321-331. ACM: New York.
- Kett, J. (2009): Die Nationalbibliografie als Linked Data – Motivation, Geschäftsmodell, Planung. Vortrag im Rahmen der SWIB 2009 (Semantic Web in Bibliotheken), Köln. http://www.swib09.de/vortraege/20091124_kett.pdf (letzter Zugriff: 29.07.2010).
- Miles, A. (2005): Quick Guide to Publishing a Thesaurus on the Semantic Web. <http://www.w3.org/TR/2005/WD-swbp-thesaurus-pubguide-20050517/> (letzter Zugriff: 29.07.2010).
- Neubert, J. (2009): Bringing the 'STW Thesaurus for Economics' onto the Web of Linked Data. Linked Data on the Web Workshop (LDOW2009), Madrid, Spanien.
- Söderbäck, A. (2009): LIBRIS - Linked Open Library Data. Vortrag im Rahmen der SWIB 2009 (Semantic Web in Bibliotheken), Köln. http://www.swib09.de/vortraege/20091124_soederbaeck.pdf (letzter Zugriff: 29.07.2010).
- Sure, Y., S. Bloehdorn, P. Haase, J. Hartmann and D. Oberle (2005): The SWRC Ontology - Semantic Web for Research Communities. 12th Portuguese Conference on Artificial Intelligence Colvilha, Portugal.